



What is Research Evidence and Why Does it Matter for Equity?

One of the most basic questions that researchers, policymakers, and the general public ask about a policy or a social program is whether it works, that is, whether it is effective in helping vulnerable families and children achieve better outcomes. In the current policy environment, there is an increasing focus on evidence-based policymaking, which uses rigorous program evaluation research evidence to guide funding decisions and future program replication or expansion.¹

What are the different types of research evidence available to measure program effectiveness?

Research evidence on program effectiveness often provides an overall “What Works” assessment of social programs based solely on the results of causal impact evaluations: studies that use experimental research designs² (i.e., random assignment) to evaluate the program’s average impacts.

- “What Works?” examines whether, on average, children or families who are offered program services (the treatment group) have better outcomes than their counterparts who are not offered program services (the control group).³

However, there is a growing call for research evidence that addresses two additional questions with important implications for policy and equity: “What Works for Whom?” and “What Works Under What Conditions?”

- “What Works for Whom?” investigates whether the program improves outcomes not only for the average participant, but also for particularly vulnerable participants (e.g., children with special needs) and/or for participants whose outcomes persistently lag behind those of their peers (e.g., black, Hispanic, and American Indian/Alaska Native children’s school readiness outcomes). This equity-related question also examines whether a program has stronger favorable impacts on subgroups of children that initially have the poorest outcomes, indicating that it may help those children catch up to their peers who start with stronger outcomes. Although assessing program effects by subgroup may present data and methodological challenges which should be addressed,⁴ understanding effects on particular subgroups is essential for programs that have the potential – or the intention – to reduce longstanding inequities in outcomes. This question can also help identify practices that could produce unintended negative impacts for certain subgroups of children.
- “What Works Under What Conditions?” recognizes that ‘on-the-ground’ implementation plays an important role in whether a program produces favorable outcomes or not, and incorporates information on how a program works in practice into program evaluation.⁵ This question encompasses equity considerations because it can uncover variation in the resources and quality of a program across locations or child subgroups.

Answering these questions requires different types of evidence. As mentioned, the first question, “What Works?” can be answered with traditional program effectiveness research evidence – that is, impact evaluations that use experimental designs to estimate the average effect of a program on participant outcomes. However, the second two questions: “What Works for Whom?” and “What Works Under What Conditions?” require an examination of the broader program context, including program logic or design and capacity. For example, if a program is not designed to provide targeted services to address the needs of a given vulnerable child subgroup, this should be taken into account when evaluating whether the program works for that subgroup. Similarly, any variation in the implementation of a program should be understood as it may result in differences in its quality or intensity, and thus potentially its effectiveness, across subgroups of children.

The diversitydatakids.org policy equity assessment of program effectiveness investigates all three of these important

diversitydatakids.org

data for a diverse and equitable future



questions using both traditional effectiveness research as well as other types of evidence and contextual program information.

Traditional “What Works” research evidence

Most projects and websites that document “what works” in social policy programs tend to draw on results from a particular type of program evaluation called an “impact evaluation.” Impact evaluations draw on experimental research designs, the most scientifically rigorous of which is a randomized controlled trial. A randomized controlled trial is a study in which participants volunteer and are randomly assigned to receive an offer of treatment (e.g. program services) or not (e.g. usual services offered in communities). This method allows researchers to assess the overall causal impact of a program on the outcomes of the average participant in the treatment group compared to the outcomes of the average participant in the control group not offered program benefits. Program impacts can also be calculated for participant subgroups.⁶

Equity-focused research evidence

Despite the merits of causal impact evaluations, judging effectiveness based only on the results of impact studies does not consider the broader program context, including how well it is designed to meet the needs of all children as well as those in the most vulnerable groups, or whether the program has adequate resources to deliver services as intended. In addition, equity-focused evaluations developed by UNICEF and other organizations systematically assess whether programs can help improve not only average outcomes but reduce inequities by targeting the most vulnerable.⁷ UNICEF has built a simulation model that provides preliminary evidence that an equity-focused service delivery approach can be more cost effective than an approach that does not prioritize improving outcomes for the most vulnerable populations.⁸ Therefore, equity-focused research evidence can provide important insights on program effectiveness that are not detected through traditional research evidence alone.

Research evidence beyond the participant population: Impacts at the population level

It is essential to evaluate both average and equity-related program impacts on the eligible or participating population when determining program effectiveness. However, diversitydatakids.org also goes one step beyond to consider research evidence at a broader level. Central to our assessment of program effectiveness is a consideration of whether a program has the capacity to reduce inequities or gaps at the population level between eligible children served by the program and the general child population. Social programs traditionally serve children from low-income families, and effectiveness research evidence usually examines whether the program works for low-income children enrolled in a program compared to their low-income counterparts not enrolled in a program. diversitydatakids.org is also interested in whether a program can help reduce inequities or gaps between low-income children and *non-low income* children at the population level. This type of research evidence has to do not only with how effective the program is in improving outcomes for low-income children, but also with the capacity of the program to serve all (or a sizable fraction of) eligible children. For example, if a hypothetical highly effective program could completely close the gap between low-income and non-low-income children but only had capacity (e.g., resources, slots) to serve five percent of low-income children, there would still remain large inequities between low-income and non-low income children at the population level. Therefore, the program would be effective for the children lucky enough to gain access, but ineffective in addressing the needs of most children due to the limited number of slots. When addressing inequities at the population level, this additional type of research evidence is a critical piece of a policy equity assessment of program effectiveness.

diversitydatakids.org
data for a diverse and equitable future



How are equity questions addressed in research evidence?

International Research Collaborations

Highly-respected international research groups have developed scientific standards and methods to guide systematic evidence reviews. Examples include the [Cochrane Collaboration](#), which focuses on systematically reviewing the results of clinical trials in the fields of public health and medicine, and the [Campbell Collaboration](#) which focuses on social policies and programs. Furthermore, international research collaboratives are increasingly developing guidelines for answering the questions of “What Works?” “for Whom?” and “Under What Conditions?” For instance, the Cochrane collaboration has developed guidelines for considering equity in effectiveness research evidence, not only for programs explicitly aimed at vulnerable groups or at reducing inequities, but for programs that may have the potential to improve equity even if it is not their primary goal.⁹ Additionally, researchers dissatisfied with traditional approaches that only consider evidence from impact evaluations have developed guidelines for conducting [realist reviews](#). These reviews incorporate information about an intervention’s implementation in different contexts into the assessment of evidence from experimental evaluations.¹⁰

U.S. Federal Level Approaches

In the United States, the types of research evidence required by the federal government to demonstrate program effectiveness have evolved across the years. In the 1990s, the Government Performance and Results Act (GPRA) focused on setting program performance goals (how a program performs in achieving defined objectives and service goals such as meeting a certain benchmark for per-pupil costs). These assessments initially focused on federal agencies but over time have shifted to individual programs with increased attention to data-driven results.¹¹ Furthermore, there is a growing consensus within some federal agencies and Congress on the importance of using impact evaluations with experimental designs¹² to draw causal inferences about whether programs successfully improve participant outcomes compared to groups that did not receive program services.¹³ Federal agencies that monitor and report on how well government programs and policies meet their objectives, such as the U.S. Government Accountability Office (GAO), assert that if programs are well developed and ready for rigorous evaluation, then conclusions about overall program effectiveness should be drawn based on the joint consideration of research evidence from performance management, program implementation and impact studies.¹⁴ While terms such as “realist review” or “equity” are not explicit in federal evaluation guidance, there is growing recognition that “What Works Under What Conditions” is an essential type of research evidence to consider when evaluating program effectiveness. For example, the authorizing legislation for the Maternal, Infant, and Early Childhood Home Visiting (MIECHV) program requires grantees to only implement programs with rigorous evidence of effectiveness, including a proven track record of program model implementation (for a minimum of three years) and adequate staff participation in training.¹⁵

The question “What Works for Whom?” is also receiving greater attention. The Congressional mandate for the Head Start Impact Study, for example, required that the study include an analysis of sources of variation in the impacts of Head Start, including variation by participant characteristics (i.e. child and family subgroups). The 2007 Head Start Act required a study on the status of Dual Language Learner children and their families participating in Head Start. These examples of subgroup research highlight the growing importance placed on understanding equity issues and how different types of children, especially the most vulnerable, are affected differently by social programs.

In conclusion, when assessing program effectiveness from an equity perspective, it is important to look within the program to understand whether it works on average, whether it works for vulnerable subgroups, and what are the contextual and implementation conditions under which it works best. This approach requires not only the use of traditional effectiveness

diversitydatakids.org

data for a diverse and equitable future



evidence but also other types of research evidence, such as information about the program's logic and capacity, including implementation. Additionally, one should look at whether the program is equipped to reduce inequities at the population level, taking into account not only its effectiveness, but also the population level of need vis-a-vis the program's resources.

Sources & notes:

- ¹ Orszag, P.R. (2010). Memorandum for the heads of executive departments and agencies: Evaluating programs for efficacy and cost-efficiency (M-10-32). Executive Office of the President, Office of Management and Budget. Retrieved from <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2010/m10-32.pdf>.
- ² Throughout this section, readers may select highlighted blue research terminology for more information on meanings and definitions.
- ³ Importantly, to maintain the integrity of the experimental design, these assessments typically evaluate the impact of an offer of program services on participant outcomes, rather than the impact of actual participation, which can lead to a biased estimate of the true impacts of the program. Many evaluations also provide a second estimate of the impact of actual program participation on the treatment group compared to the control group, sometimes adjusting for the potential sources of bias.
- ⁴ For more information on potential data and methodological challenges inherent in subgroup analysis, see: *Interagency meeting on subgroup analysis*. Child Care & Early Education, Research Connections. Retrieved from <http://www.researchconnections.org/childcare/datamethods/interagency-meeting.jsp>; and Bloom, H.S. & Michalopoulos, C. (2010). *When is the story in the subgroups? Strategies for interpreting and reporting intervention effects for subgroups*. New York, NY: MDRC. Retrieved from <http://www.mdrc.org/publications/551/full.pdf>.
- ⁵ Gueron, J.M. & Rolston, H. (2013). *Fighting for reliable evidence*. New York, N.Y., Russell Sage Foundation.
- ⁶ For more information on impact evaluation review methods, see Higgins, J.P.T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions, Version 5.1.0*. The Cochrane Collaboration. Retrieved from <http://handbook.cochrane.org/>.
- ⁷ Bamberger, M. & Segone, M. (2011). *How to design and manage equity-focused evaluations*. UNICEF Evaluation Office. Retrieved from http://mymande.org/sites/default/files/EWPS_Equity_focused_evaluations.pdf; Campbell & Cochrane Equity Methods Group. Retrieved from <http://equity.cochrane.org/>.
- ⁸ Carrera, C., Azrack, A., Begkoyian, G., Pfaffmann, J., Ribaira, E., O'Connell, T.,..., Knipperberg, R. (2012). The comparative cost-effectiveness of an equity-focused approach to child survival, health, and nutrition: A modelling approach. *The Lancet*, 380(9850), 1341–1351. doi:10.1016/S0140-6736(12)61378-6. Retrieved from http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2812%2961378-6/fulltext#article_upsell.
- ⁹ Welch, V., Petticrew, M., Tugwell, P., Moher, D., O'Neill, J., Waters, E., White, H., and the PRISMA-Equity Bellagio Group. (2012). PRISMA-Equity 2012 extension: Reporting guidelines for systematic reviews with a focus on health equity. *PLoS Med* 9(10): e1001333. doi:10.1371/journal.pmed.1001333. Retrieved from <http://www.plosmedicine.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pmed.1001333&representation=PDF>.
- ¹⁰ Greenhalgh, T., Kristjansson, E., & Robinson, V. (2007). Realist review to understand the efficacy of school feeding programmes. *BMJ*, 335, 858-861. Retrieved from http://r4d.dfid.gov.uk/PDF/Articles/SR_SchoolFeeding_Analysis.pdf.
- ¹¹ Heinrich, C.J. (2011). How credible is the evidence, and does it matter? An analysis of the Program Assessment Rating Tool. *Public Administration Review*, 72(1), 123-134. doi: 10.1111/j.1540-6210.2011.02490.x. Retrieved from https://www.utexas.edu/lbj/sites/default/files/file/Heinrich_PAR%202012.pdf; Lew, J.J. & Zients, J. (2011). Memorandum for the heads of executive departments and agencies: Delivering on the Accountable Government Initiative and implementing the CPRA Modernization Act of 2010 (M-11-17). Executive Office of the President, Office of Management and Budget. Retrieved from <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2011/m11-17.pdf>
- ¹² An impact evaluation is a systematic study measuring whether a program successfully impacts participant outcomes compared to a randomly assigned group who did not receive the program's services.
- ¹³ Gueron, (2013), op. cit.
- ¹⁴ For more information about the differences between program performance and program evaluation, see United States Government Accountability Office. (2011). *Performance measurement and evaluation: Definitions and relationships* (GAO-11-646SP). Retrieved from <http://www.gao.gov/assets/80/77277.pdf>.
- ¹⁵ Paulsell, D., Avellar, S., Sama Martin, E., & Del Grosso, P. (2011). *Home Visiting Evidence of Effectiveness Review: Executive Summary*. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Washington, DC.